AD_____

Award Number: DAMD17-02-1-0597

TITLE: Retroelements and Genetic Instability in Breast Cancer

PRINCIPAL INVESTIGATOR: Victoria Perepelitsa Belancio
Prescott L. Deininger

CONTRACTING ORGANIZATION: Tulane University Health Sciences Center
New Orleans, Louisiana 70112

REPORT DATE: April 2004

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 2004	Annual Summary (1 Apr 03-31 Mar 04)

4. TITLE AND SUBTITLE

Retroelements and Genetic Instability in Breast Cancer

5. FUNDING NUMBERS

DAMD17-02-1-0597

6. AUTHOR(S)

Victoria Perepelitsa Belancio
Prescott L. Deininger

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Tulane University Health Sciences Center
New Orleans, Louisiana 70112

E-Mail: vperepe@tulane.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

LINE1 is a mammalian retroelement that contributes to genomic instability. The full extent of LINE-1 mobility in somatic tissues and tumors is not known. L1 expression is extremely low in differentiated cells except for testis, but it is significantly elevated in breast malignancies. This suggests that posttranscriptional mechanisms are involved in limitation of L1 expression. We demonstrated that the use of the polyA sites located within the L1.3 genome limits the amount of full-length L1.3 mRNAs. L1 polyA signals can be functional when fragments of L1.3 element are inserted into 3' UTRs of genes. This unique attenuation mechanism helps to minimize the rate of L1 retrotransposition, but may also increase the negative impact of these insertion events on the genome after their insertion. Human EST database searches suggest that the polyA signals may also play a role in regulation of L1 expression in a tissue and/or tumor specific manner with breast cancer tissues supporting the least efficient L1 polyadenylation. The EST data are strengthened by significant differences in the L1 RNA profiles between transiently transfected breast cancer and nonbreast cancer cell lines. These observations suggest a potential global change in the mechanism of polyadenylation process upon malignant transformation of mammary gland.

14. SUBJECT TERMS

Regulation of gene expression, LINE1, polyadenylation, genomic instability, tissue specific expression.

15. NUMBER OF PAGES

22

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Introduction.

LINE1 is the most abundant active human retroelement, and it contributes to genomic instability. The full extent of LINE-1 mobility in somatic tissues and tumors is not known. L1 expression is extremely low in differentiated cells except for testis, but it is significantly elevated in breast malignancies. These observations suggest that posttranscriptional mechanisms are involved in limitation of L1 expression. We demonstrated that the use of the polyA sites located within the L1.3 genome limits the amount of full-length L1.3 mRNAs present in mammalian cells. Internal L1 polyA signals can also be functional when fragments of L1.3 elements are inserted into 3' untranslated regions of mammalian genes. This unique attenuation mechanism helps to minimize the rate of L1 retrotransposition, but may also increase the negative impact of these insertion events on the genome after their insertion. Human EST database searches suggest that these internal polyA signals may also play a role in regulation of L1 expression in a tissue and/or tumor specific manner with breast cancer tissues supporting the least efficient internal L1 polyadenylation. The EST data are strengthened by significant differences in the L1 RNA profiles between breast cancer and nonbreast cancer cell lines detected by northern blotting of transiently transfected human L1.3. These observations suggest a potential global change in the mechanism of polyadenylation process upon malignant transformation of mammary gland.

Body.

According to the approved Statement of Work we have made the following progress in the second year of funding:

Task 1. To identify functional polyA sites in the L1.3 genome.

A-B. We developed a sensitive northern blotting assay that allowed detection of truncated L1 related RNA species produced from a construct containing a full-length L1.3. The result was consistent with premature internal polyadenylation. We confirmed that multiple predicted canonical and noncanonical internal polyA signals identified within L1.3 coding region are functional by site-directed mutagenesis, 3'RACE, and human and mouse EST database analysis (Perepelitsa-Belancio and Deininger, Nature Genetics, 2003, reprint of the publication is attached).

The experimental design described in this portion of the original proposal utilized a reporter system with an intron located upstream of the *Renilla* luciferase (Rluc) gene (Fig.1). We created a series of constructs containing fragments of the L1.3 genome with the strongest predicted polyA signals in the above-described intron (opiluc constructs). As stated in the previous Annual Summary Report, we rendered this approach unfit for our study due to the competition between splicing and polyadenylation (1). We subsequently developed a method for detection of functional polyadenylation signals within the entire L1.3 genome (2). Discovery of functional polyA sites located throughout the L1.3 coding region led us to revisit this reporter gene system. It can address an important question relevant to the impact that L1 insertions have on the mammalian genomes. The question is: can internal L1 polyA signals be utilized upon element's insertion into an intron of a mammalian gene and attenuate its expression? In our previous work, in order to detecte functional polyA sites located in the intron of the reporter gene, we relied on reduction in the Rluc activity. However, if the polyA sites are used inefficiently due to the competition with splicing, the changes in the enzymatic activity may fall within experimental error of the assay. Therefore such subtle differences would not be detectable in the transient transfections, yet relevant in the adequate biological systems. For example, L1 insertions into the introns and 3' untranslated regions (UTR) of mammalian genes are less detrimental to the genomes. As a result, a number of L1 fragments are sometimes found within introns. Therefore, for a gene containing multiple introns (which is often the case for mammalian genes) a combined impact of premature termination may significantly impair protein production. We are currently working out the conditions for a 3' RACE analysis of RNA species produced by opiluc constructs upon transient transfections in NIH 3T3 cells.

C-F. We previously reported that insertions of fragments of L1.3 genome into the 3' UTR of a reporter gene (FRM/N series of constructs, Fig. 2) significantly decreased the total amount of produced mRNA (Fig. 3 page 14 of appendices Annual Summary Report 2003). This result was inconclusive for identifying whether putative L1.3 polyA sites were functional because (i) no slower migrating bands were detected on the northern blots and (ii) reduction of total mRNA could have resulted from decreased stability of the mRNAs

containing L1.3 fragments. Alternatively, the result may be a combination of the two. At the time we were interested in developing an adequate approach to detecting functional polyA signals in the system that would be the most biologically relevant to the L1.3 life cycle (see above). However, with the knowledge we gained by identifying functional internal L1.3 polyA sites in the context of the full-length L1.3 element, we were compelled to revisit the FRM/N system. This system can provide an answer to the important question of whether L1 polyA signals can be utilized when L1 fragments are inserted into the 3' UTRs of mammalian genes, a phenomenon often found in mammalian genomes. We used a sensitive 3' RACE analysis to detect any truncated RNA species resulted from utilization of the internal L1.3 polyA signals from FRM/N our control vector and FRM/N#2, a construct that contains a 1 kb. L1.3 fragment cloned into the 3'UTR of FRM/N (Fig. 3). The choice was dictated by the observation that northern blotting assay may not be sensitive enough to detect truncated products. As expected, we detected multiple truncated bands (250 bp. to 1.2 kb.) that in size corresponded to the positions of the polyA signals in L1.3 fragment. Sequence analysis of some of the truncated bands revealed the usage of noncanonical L1.3 polyA sites. This result also provides a direct explanation of the observation that majority of the L1 insertions identified in the human and mouse genes are in the reverse orientation (3).

We also removed a very strong SV40 polyadenylation signal present in the FRM/N and FRM/N#2 vectors. This approach is designed to test whether the utilization of internal L1.3 polyadenylation sites varies depending on the strength of the polyA signal at the end of a gene. The rational for this experiment is based on the reported competition between polyadenylation signals depending on their relative strength and position (). We are currently working out optimal conditions for the 3'RACE analysis of FRM/NΔSV40 and FRM/N#2ΔSV40 vectors. We do not expect to obtain any quantitative data from this experiment for 3'RACE is not a quantitative assay. However, it will give us a good understanding of the influence of genomic sequences surrounding an L1 fragment on the utilization of the polyA sites it encodes. We expect that there may be a difference in the band pattern produced from FRM/N#2ΔSV40 and FRM/N#2 vectors.

The data produced by the experiments described in A-B and C-F will be combined for a publication with the bioinformatic data demonstrating the frequency of the L1 insertions in the introns and 3' UTRs of human genes.

Task 2. To identify the function and relative strength of functional polyA sites in normal and breast cancer cells.

We have compared the L1.3 RNA profiles in mouse fibroblast NIH 3T3 cell line and in aggressive human breast cancer cell line, SK-Br-3, that overexpresses ErbB3 receptor. Northern blot analysis with the strand-specific 5'-UTR probe demonstrated significant differences in the utilization of the internal polyadenylation sites between these two cell lines (Fig. 4). SK-Br-3 breast cancer cell line does not support efficient internal polyA site usage making the full-length L1.3 mRNA be the major product in this cell type. In contrast, only 2-4% of the total L1 related products correspond to the full-length L1, when the same construct is transiently transfected in NIH 3T3, HeLa, or chicken fibroblasts. This result strongly indicates that there is a significant difference in the usage

of the internal polyA signals during L1 transcription in breast cancer vs. fibroblast cell lines.

These data are consistent with our preliminary bioinformatic results. We used 100 bp. fragments of L1.3 as BLAST query sequences to search existing human EST databases. The rationale behind this approach was to determine whether there is a tissue- and/or cancer-specific difference in the usage of internal polyA sites by active endogenous human L1 elements. We found out that proportion of L1 related 3'end ESTs varied significantly among different human tissues (Table 1). The distribution of the L1 3'ends in lung, stomach, and retina were consistent with the L1 profiles detected by northern blot analysis in mouse fibroblast. On the contrary, the pattern of the L1 related 3' ends in placenta, testes, ovary, and brain differed from that detected in other tissues and by northern blotting in NIH 3T3 cells. We also tested whether there is a variation in the polyA site usage between normal and cancer tissues of different human organs (Table 2). We detected significant difference in the proportions of the L1 related 3'end ESTs in normal vs. tumor tissues of brain and ovary. Although we found no normal breast L1 sequences for comparison, the majority of L1 elements found in breast cancer were full-length. Additionally, the proportion of ESTs corresponding to the full-length L1 element differed significantly, 4 to 7 fold, between some normal and tumor tissues supporting a potential increase in L1 retrotransposition rate upon malignant transformation The EST data from mammary gland, even though very limited, correlates with the profile of L1 internal polyadenylation detected by northern blotting in SK-Br-3 breast cancer cell line. More detailed analysis of the human and mouse EST database will be required to strengthen these preliminary results. Unfortunately, this kind of analysis is beyond our technical capability. Our data suggest that internal L1 polyA sites are differentially utilized in human tissues and upon malignant transformation of breast, brain, and ovary. The data indicate that internal polyadenylation may play a role in regulation of the amount of the full-length L1 element and therefore retrotransposition.

We will continue to explore L1 polyadenylation profiles in other breast cancer cell lines.

We will also expend the experiments described in Task 1 (C-F) by using FRM/N-based constructs in different breast cancer cell lines.

Task 3. To identify the effect of functional polyA sites on the rate of L1 retrotransposition.

We have shown that mutation of functional internal L1.3 polyA signals results in the more efficient utilization of the polyA sites located in the vicinity (Perepelitsa-Belancio and Deininger, 2003). Removal of five functional L1.3 internal polyadenylation signals lead to a modest but statistically significant increase in the rate of retrotransposition in HeLa cells. We will next determine the rate of L1.3 retrotransposition in SK-Br-3 cell line and other breast cancer cell lines.

Key Research Accomplishments.

Task 1. To identify functional polyA sites in L1.3 genome (Months 1-25)
Task 2. To identify the function and relative strength of functional polyA sites in normal and breast cancer cells (Months 25-29).
Task 3a. To identify the effect of functional polyA sites on the rate of L1 retrotransposition (Months 29-36)

- We determined that putative canonical and noncanonical polyA sites found in the coding region of the human L1.3 element are functional.
- We demonstrated that removal of five functional polyA signals resulted in a statistically significant increase in the rate of L1 retrotransposition in HeLa cells.
- We demonstrate by northern blots of transiently transfected breast cancer cells that there is a much less efficient utilization of the internal L1 polyA signals in SK-Br-3 breast cancer cell line.
- We performed human EST database search to determine whether there are any differences in the usage of internal polyA signal in breast tissues.
- We performed human EST database search to determine whether there are any differences in the usage of internal polyA signal between normal human tissues human breast cancers.
- We performed a 3' RACE analysis to determine functional polyA signals in L1 fragment located in the 3' UTR of the reporter gene.

Reportable outcomes:

1. An appended manuscript "RNA truncation by premature polyadenylation attenuates human mobile element activity" has been published in journal of *Nature Genetics* Volume 35, Number 4, December 2003.

2. Responses to the publication: Press release "The silence of the LINEs", "Silent garbage"

3. Appended is an abstract "RNA truncation by premature polyadenylation attenuates human mobile element activity" of the presentation that took place at the annual Molecular and Cellular biology Program Retreat, Tulane University, New Orleans, LA in October, 2003.

4. A Ph.D. in Molecular and Cellular Biology from Tulane University

5. 17th Morris F. Shaffer and Margaret H.D. Smith-Shaffer Award for Excellence in Research, received in May, 2004.

6. Patent application 60/445,945 (extension filed on February 7, 2004).
Deininger, Prescott L. and Victoria Perepelitsa Belancio
Entitled: Mammalian Retrotransposable Elements.

Conclusions:

1. We demonstrated that both canonical and noncanonical polyA sites found in the coding region of the human L1.3 element are functional, and they attenuate the amount of the full-length mRNA and the rate of retrotransposition in HeLa cells.

2. Our EST data and transient transfections of L1.3 expression cassette of human breast cancer cell line suggest that the process of internal polyadenylation may be involved in regulation of the amount of the full-length L1 mRNA in a tissue specific manner and upon malignant transformation.

3. Our 3' RACE analysis demonstrated that internal L1 polyadenylation sites are used during transcription of a luciferase reporter gene when the L1 fragment is cloned into the 3' UTR of the gene. This finding has significant implications for our understanding of mammalian gene expression.

Reference List

1. Peterson,M.L., Bryman,M.B., Peiter,M. and Cowan,C. (1994) Exon size affects competition between splicing and cleavage- polyadenylation in the immunoglobulin mu gene. *Mol.Cell Biol.*, **14**, 77-86.

2. Perepelitsa-Belancio V. and Deininger P. (2003) RNA truncation by premature polyadenylation attenuates human mobile element activity. *Nat.Genet.*, **35**.

3. Medstrand,P., van de Lagemaat,L.N. and Mager,D.L. (2002) Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.*, **12**, 1483-1495.

Appendices.

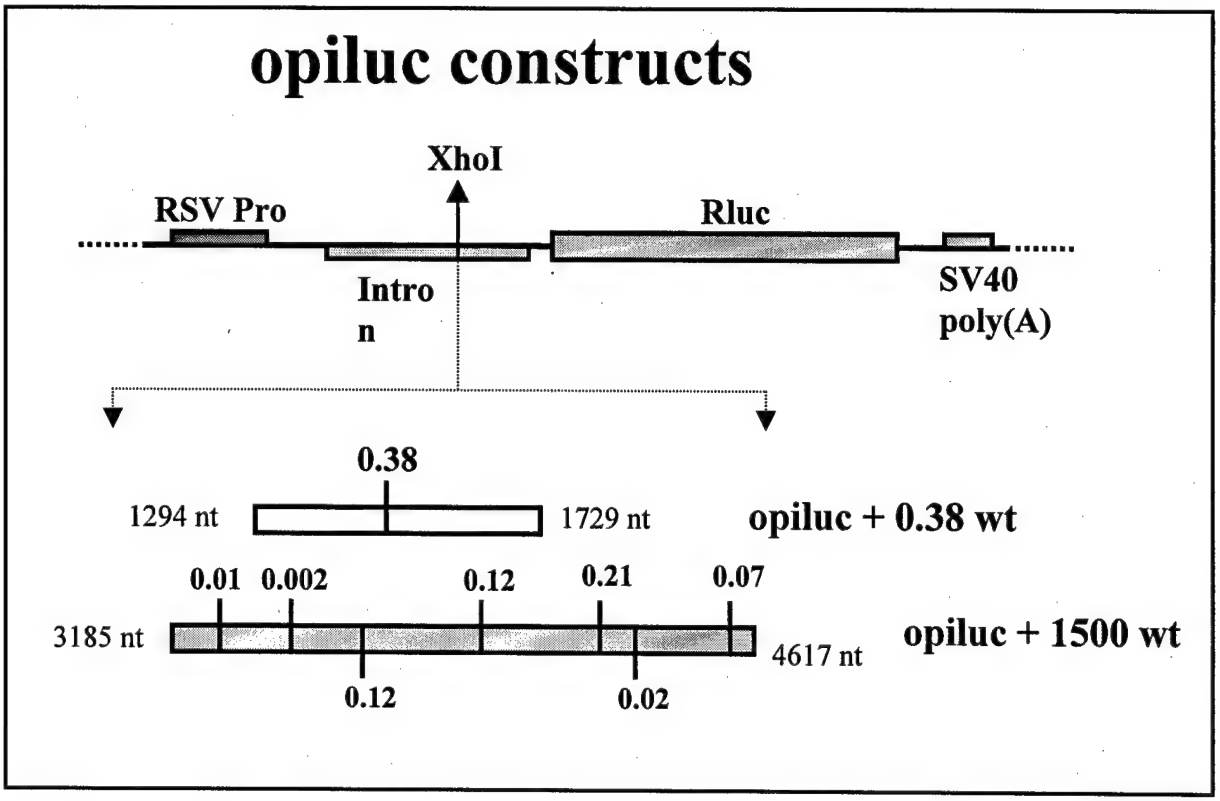

Figure 1. The opiluc reporter construct. We have built this reporter vector, which expresses the *Renilla* luciferase gene from the RSV promoter. The vector has a multilinker in an intron region labeled XhoI which is the primary site used for cloning of the polyA sites. Below it is the segment of the L1.3 sequence that we have subcloned. The vertical lines with numbers represent predications of the relative strengths of the poly A sites using the polyadq program.

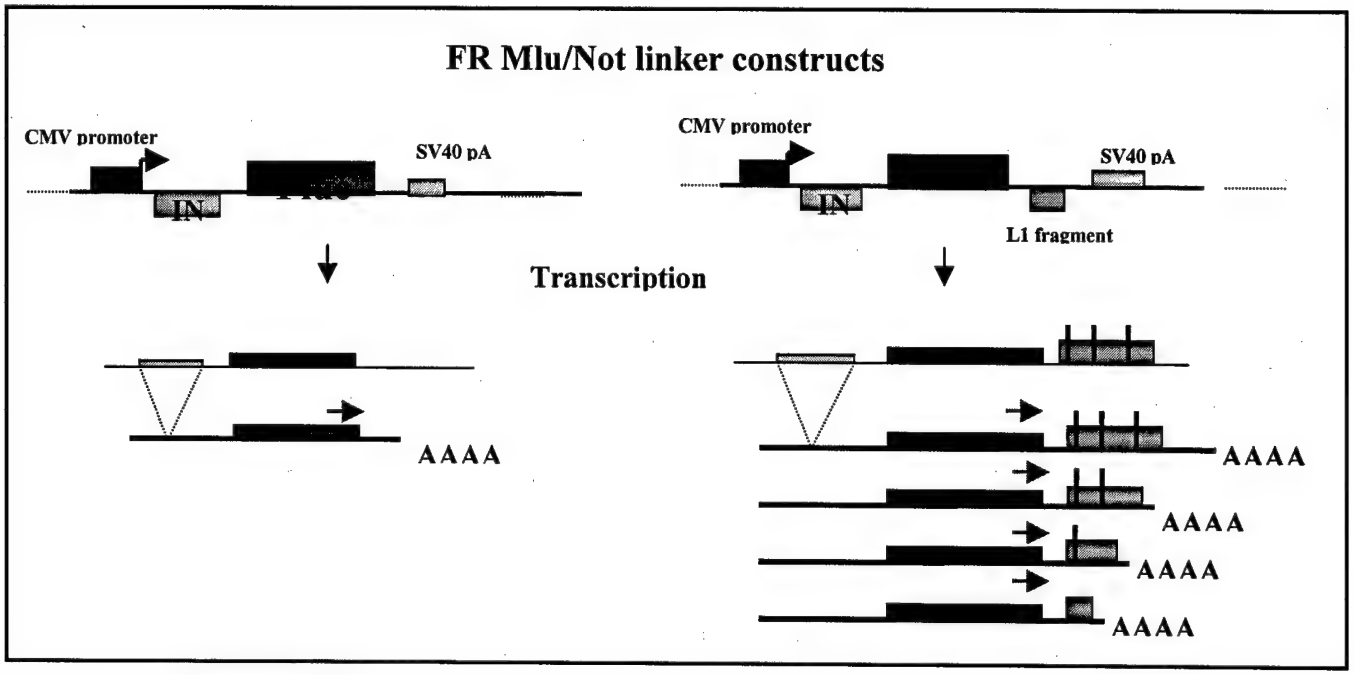

Figure 2. FR Mlu/Not linker constructs.1 kb fragments of the L1.3 genome were cloned into the 3' UTR of the *Firefly*(F) luciferase reporter gene. Vertical black lines represent functional polyA signals identified in the L1.3 genome. FR M/N construct and its mRNA products are shown on the left side of the panel. FRM/N#2 construct and its potential mRNA products (if the L1.3 polyA sites are used during transcription) are shown on the right side of the panel. A red arrow indicates the position of the upstream Fluciferase-specific primer used in the 3'RACE analysis.

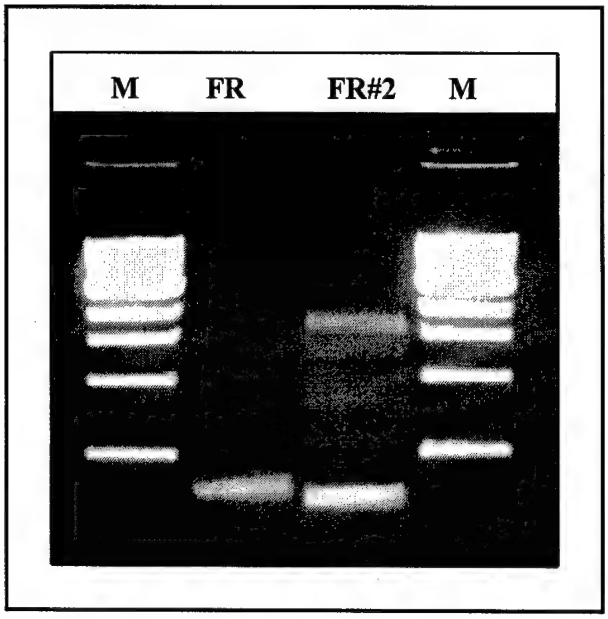

Figure 3. 3'RACE analysis of the FR and FR#2 constructs with the Firefly luciferase specific primer. NIH 3T3 cells were transfected with FR or FR#2 expression cassettes. Total RNA was harvested 24 h. post transfection, polyA selected and subjected to the 3'RACE analysis. The size of the band detected for FR construct is consistent with the usage of the SV40 polyadenylation signal. This was also confirmed by sequence analysis of the DNA Multiple bands observed for FR#2 construct are consistent with the usage of polyA sites located in the L1.3 fragment cloned in the 3' UTR of the luciferase reporter gene.

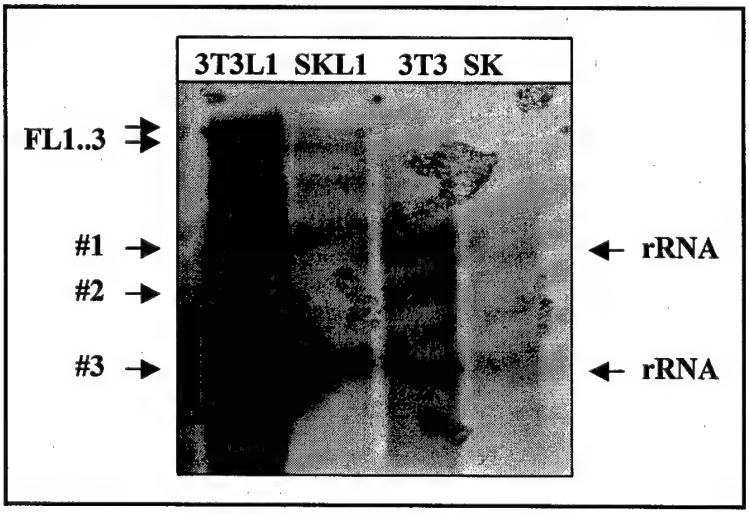

Figure 4. Profiles of wt L1 internal polyadenylation in mouse fibroblasts and human breast cancer cell lines. NIH 3T3 and SK-Br-3 cells were transiently transfected with the wt L1.3 expression cassette. Total RNA was harvested 24 hr. post transfection, polyA selected, fractionated by electrophoresis, and transferred onto the nylon membrane. Northern blot analysis was carried out with the strand-specific 5'UTR probe. 3T3L1 and SKL1 are NIH 3T3 and SK-Br-3 cells transfected with the L1.3 expression cassette, 3T3 and SK are nontransfected NIH 3T3 and SK-Br-3 cells respectively.

	Lung	Testis	Stomach	Retina	Ovary	Placenta	Brain	Breast
1-100	3.1	28.3	2.5	4.1	2.9	6.4	5.0	0.0
500-600	2.2	1.9	1.2	2.6	1.5	0.0	0.8	0.0
1000-1100	6.6	1.9	14.8	10.8	10.3		8.3	0.0
1500-1600		1.9				8.5	13.2	0.0
2000-2100	5.9	7.5	16.0	12.3	8.8	8.5	3.3	0.0
2500-2600	3.4	3.8	1.2	1.5	0.0	2.1	4.1	0.0
3000-3100	8.4		1.2	3.1	4.4	17.0		0.0
3500-3600	8.1	7.5	1.2	3.1	11.8	17.0	4.1	0.0
4000-4100	3.4	5.7	1.2	1.5	7.4	2.1	4.1	0.0
4500-4600	5.6	1.9	4.9	3.1	7.4	2.1	2.5	0.0
5000-5100	9.4	7.5	4.9	3.6	5.9	2.1	8.3	
5500-5600	10.0	7.5	1.2	3.6	4.4	2.1	8.3	0.0
6000-6059	22.2	15.1	14.8	5.1	19.1	12.8	21.5	60.0
# of ESTs	320	53	81	195	68	47	121	5

Table 1. Percentage of human L1 ESTs in different tissues. ESTs were identified using BLAST and the L1 consensus query region shown in the left column. Fractions of ESTs identified in the various bins are shown, along with the total number of transcripts studied for each tissue type. The highest fraction for each tissue is highlighted in red, with those of secondary abundance shown in green. Yellow marks the proportion of full-length transcripts.

	Lung		Brain		Ovary		Breast	
	N	T	N	T	N	T	N	T
1-100	3.4	2.4	2.7	8.3	4	2.3		0.0
500-600	2.5	1.2	1.4	0	0	2.3		0.0
1000-1100	8	2.4	4.1		12	9.3		0.0
1500-1600		4.9	17.8	6.2		2.3		0.0
2000-2100	5	8.5	2.7	4.2	4	11.6		0.0
2500-2600	3.4	3.7	6.8	0	0	0		0.0
3000-3100	8.4	8.5		6.2	4	4.6		0.0
3500-3600	5		5.5	2.1	8			0.0
4000-4100	3.4	3.7	5.5	2.1	8	7		0.0
4500-4600	4.2	9.8	4.1	0	8	7		0.0
5000-5100	8.4	12.2	4.1		4	7		
5500-5600	12.6	2.4	11	4.2	4	4.7		0.0
6000-6059	21.9	23.2	11	37.5	4	27.9		60.0
# of ESTs	238	82	73	48	25	43	0	5

Table 2. Percentage of human L1 ESTs in normal vs. tumor tissues. ESTs were identified using BLAST and the L1 consensus query region shown in the left column. Fractions of ESTs identified in the various bins are shown, along with the total number of transcripts studied for each tissue type. The highest fraction for each tissue is highlighted in red, with those of secondary abundance shown in green. Yellow marks the proportion of full-length transcripts.

RNA TRUNCATION BY PREMATURE POLYADENYLATION ATTENUATES HUMAN MOBILE ELEMENT ACTIVITY.

Victoria P. Belancio and Prescott Deininger

LINE1 is the most prevalent human retroelement, and it contributes to genomic instability. LINE1 elements comprise 17% of the human genome, which translates into about $5x10^5$ L1 copies, the majority of which are truncated at their 5' end. The full extent of LINE-1 mobility in somatic tissues and particularly in tumors is still not known. LINE-1 is the only member of the autonomous non-LTR retrotransposones family that is currently active. It is believed that about 60 active copies of full-length LINE-1 are present in the human genome.

L1 expression is extremely low in all cell types of a mature organism except for testis. In contrast, significantly higher levels of LINE-1 expression were found in various cancer cells. The known factors involved in regulation of LINE1 expression, such as hypomethylation often associated with malignant transformation and promoter activity, cannot fully explain the observed pattern of expression. In addition, full-length LINE-1 is not detected by Northern blotting in cell culture even when transiently expressed from the CMV promoter. Together, these observations suggest that posttranscriptional mechanisms might be involved in regulation or limitation of L1 expression. By using a polyadq program, we identified 20 putative polyadenylation (polyA) sites located only in the sense strand of the L1.3 genome. We hypothesize that the use of the putative polyA sites located within the L1.3 genome and RNA instability limit the amount of full-length L1.3 mRNAs present in mammalian cells.

To address this question we developed a sensitive northern blot assay that allows detection of the full-length L1 mRNA as well as any RNA species produced through internal polyadenylation. We determined that the A-rich coding strand of the L1 elements contains numerous internal polyadenylation sites that attenuate full-length L1 RNA formation by about 50 fold. There is tremendous redundancy in these internal polyadenylation sites, and their presence is conserved throughout mammalian L1 elements. This unique attenuation mechanism helps to minimize the rate of L1 retrotransposition, but may also increase the negative impact of these insertion events on the genome.

nature genetics

RNA truncation by premature polyadenylation attenuates human mobile element activity

Victoria Perepelitsa-Belancio & Prescott Deininger

Long interspersed elements (LINE-1s, also called L1s) are the only active members of the autonomous, non–long terminal repeat (LTR) retrotransposon family, which reshapes mammalian genomes in many different ways[1–5]. LINE-1 expression is low in most differentiated cells but high in some cancer cells, in testis and during embryonic development[6–12]. To minimize the negative impact on their hosts' genomes, many mobile elements strategically limit their amplification potential, particularly in somatic cells[13–15]. Here we show that the A-rich coding strand of the human LINE-1 contains multiple functional canonical and noncanonical polyadenylation (poly(A)) signals, resulting in truncation of full-length transcripts by premature polyadenylation. This attenuation lowers the rate of retrotransposition in assays using HeLa cells. It probably also increases the negative effects of LINE-1 insertions into genes[16].

Analysis of the LINE-1.3 sequence with the POLYADQ[17] program identified 19 potential poly(A) signals, AATAAA and ATTAAA, in the sense orientation but only 2 in the antisense strand of the human LINE-1 element (**Supplementary Table 1** online). Several sites are predicted to be much stronger than the relatively weak poly(A) site found at the 3′ end of the LINE-1 element[18]. Gorilla and mouse elements showed a similar enrichment in the sense strand (**Fig. 1**). Additionally, there are 141 noncanonical poly(A) sites distributed throughout the LINE-1.3 genome, deviating from the consensus by only one base (**Supplementary Table 1** online). The widespread presence of these poly(A) signals suggests that they have a conserved function, perhaps limiting, or regulating, LINE-1 retrotransposition.

By transiently expressing an active human LINE-1.3 (ref. 19) element in mouse NIH 3T3 cells, we detected two poly(A), high-molecular-weight, LINE-1.3-specific bands that migrated between 7.4 and 9.4 kb (**Fig. 2a**). The doublet was observed with probes to both the 5′ UTR and the neomycin marker at the 3′ end of the LINE-1.3 construct (**Fig. 2b**), indicating that these RNAs represent full-length transcripts. The presence of the doublet is consistent with the inefficient splicing of the intron[20] (**Fig. 2b**). Multiple, faster-migrating species were also present and much more abundant than the full-length RNAs. The sizes of these transcripts roughly corresponded to the positions expected with use of the putative poly(A) sites identified in the LINE-1.3 sequence (**Fig. 2a**). The most intense band, band 3, correlated with the strongest predicted poly(A) site. The bands were absent from the flow-through fraction of the oligo-dT selection and were not detected by the neomycin strand-specific probe located downstream of these termination sites (data not shown).

In NIH3T3 cells transfected with the LINE-1.3 expression vector lacking the intron-containing *neo*^{r} tag, the pattern of the truncated bands was identical to that detected with the 5′ UTR probe (**Fig. 2a**). Although there was a higher proportion of full-length transcripts, suggesting that the *neo*^{r} tag interferes somewhat with full-length RNA production, these data indicate that the premature polyadenylation is not an artifact of the splicing introduced in the reporter system. We therefore used the LINE-1.3*neo*^{r} expression vector for future experiments to correlate RNA profiles with the rate of retrotransposition.

Northern blotting of RNAs from NIH3T3 cells transfected with a mouse element, LINE-1spa[21,22] (**Fig. 2a**), detected abundant truncated bands corresponding in size to the positions of the putative poly(A) signals in this element. This shows that premature polyadenylation is conserved in mammalian LINE-1 elements. The presence of these polyadenylation sites correlates with the high (~40%) A-residue content in the LINE-1 coding region.

Two inactivating point mutations introduced into the hexamer of the strongest predicted poly(A) site[23] (mutant 1) resulted in the complete loss of the wild-type band 3 in NIH3T3 cells (**Fig. 2a**) and more efficient use of the nearby poly(A) sites (**Fig. 2a**). The amount of full-length LINE-1.3 mRNA (**Fig. 2a**) was similar in the wild type and mutant 1 (**Supplementary Table 2** online). The RNA profile of LINE-1.3 with mutations in canonical poly(A) sites that were expected to produce bands 4 and 5 in the background of mutant 1 (mutant 3) was similar to the band pattern of mutant 1 (**Supplementary Fig. 1** online). This suggested that use of some of the multiple noncanonical poly(A) signals in the region compensates for the removal of the classical sites. Inactivation of the noncanonical poly(A) signals at positions 2,053 or 2,079 in the background of mutant 3 did not significantly change the intensity of band 4 (**Supplementary Fig. 1** online and data not shown). But band 4 disappeared in the mutant lacking both of these poly(A) sites in the mutant 3 background (mutant 5) (**Fig. 2a**). Removal of these poly(A) signals shifted the RNA intensities to higher-molecular-weight bands, indicating more efficient use of the downstream sites, as judged by the intensity of band 2 (**Fig. 2a** and **Supplementary Table 2** online). Thus, internal poly(A) sites present in the LINE-1.3 genome

Tulane Cancer Center, SL66, and Department of Environmental Health Sciences, Tulane University Health Sciences Center, 1430 Tulane Ave., New Orleans, Louisiana 70112, USA. Correspondence should be addressed to P.D. (pdeinin@tulane.edu).

Published online 16 November 2003; doi:10.1038/ng1269

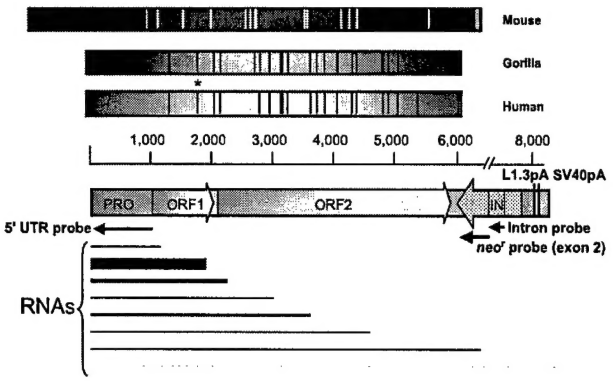

Figure 1 Putative poly(A) sites in the human (LINE1.3), gorilla (LINE-1Gg-1A) and mouse (LINE-1spa) LINE-1 elements. The genomes were aligned according to the beginning of the open reading frame (ORF) 1. An asterisk marks the strongest predicted poly(A) site. Below the nucleotide scale is a schematic of the LINE-1 retrotransposition cassette[29]. The approximate positions of the endogenous LINE-1 promoter (PRO), coding regions (ORF1 and ORF2), 3′ UTR and poly(A) site (L1.3pA), intron (IN) and inverted neomycin resistance gene (large arrow pointing to the left containing the intron sequence) are shown. The SV40 polyadenylation site (SV40pA) is located immediately downstream of the LINE-1 cassette. Horizontal lines represent the predicted mRNA species with the thickness of the line suggesting that different sites probably truncate higher proportions of the RNAs. The bottom dotted line is consistent with read-through transcripts[18]. Thick black horizontal arrows labeled 5′ UTR probe, *neo*^{r} probe and intron probe reflect genomic positions of the strand-specific probes used for northern-blot analysis.

may have a modest effect on limiting the LINE-1.3 expression individually but together result in substantial transcriptional attenuation.

Despite limitations in detection of modest changes of full-length LINE-1 RNA levels, mutants 1 and 5 both caused a statistically significant increase in retrotransposition rate, by a factor of almost 2, in HeLa cells (**Supplementary Table 3** and **Supplementary Methods** online).

Endogenous LINE-1 RNAs from human Ntera2 and HeLa cells had a similar series of truncated transcripts (**Fig. 2c**). We also compared the band distribution in transiently transfected and nontransfected HeLa cells to those in transfected chicken and mouse cells (**Fig. 2c**). We observed similar patterns with only modest differences, suggesting species-specific or cell type–specific variations in use of poly(A) sites. Our results are consistent with a number of previous observations that both human and mouse cells contain a number of smaller, heterogeneous, LINE-1-related bands[6,12,24,25] of unknown origin, with only a few cell types containing full-length LINE-1 RNA[6,12,25]. Full-length transcripts, but not truncated bands, were detected in cytoplasmic, poly(A)$^+$-selected RNAs from Ntera2 cells[25]. Therefore, we tested cytoplasmic versus nuclear poly(A)-selected fractions in transfected NIH3T3 cells and found that the truncated bands, as well as the full-length transcripts, were almost exclusively nuclear (**Supplementary Fig. 2** online).This is consistent with observations that truncated bands from mouse LINE-1 are primarily in the nucleus[12].

We found that most of the transcripts in the human expressed-sequence tag (EST) database with 3′ end sequencing of LINE-1 elements are truncated at positions predicted by the internal poly(A) sites (**Fig. 3**), further confirming the use of the internal poly(A) sites in the endogenous LINE-1 elements. A similar analysis of mouse EST

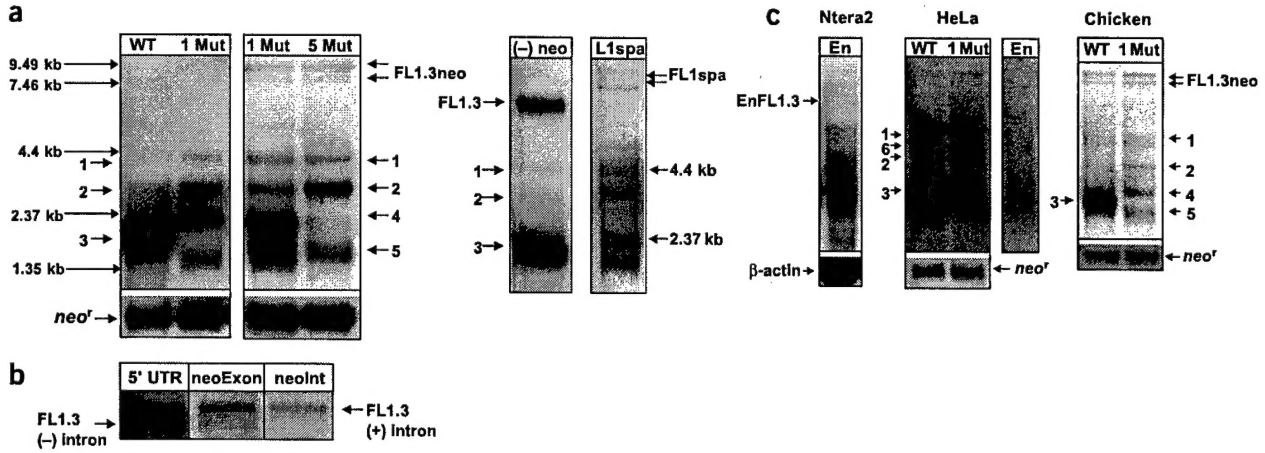

Figure 2 Use of the internal poly(A) sites of LINE-1 element. (**a**) Northern-blot analysis of LINE-1.3 RNA species produced by the wild-type (WT) human element, an element with a mutation in the strongest predicted poly(A) site (mutant 1; 1 Mut), a LINE-1.3 element lacking five poly(A) sites (mutant 5; 5 Mut), wild-type LINE-1.3 element without the *neo*^{r} tag (– neo) and wild-type mouse LINE-1spa (L1spa) in NIH 3T3 cells probed with the human or mouse strand-specific 5′ UTR probe. FL1.3 is full-length human LINE-1.3 mRNA (the size difference of FL1.3 in – neo lane is due to the absence of *neo*^{r} in the expression cassette). FL1spa is full-length mouse LINE-1spa mRNA. *neo*^{r} is neomycin-resistance mRNA detected with the randomly labeled probe to exon 2 of the *neo*^{r} gene. Long horizontal arrows correspond to the positions of the molecular weight RNA marker (Invitrogen). RNA bands that are described specifically in the text are numbered. (**b**) The high-molecular-weight doublet is due to inefficient splicing. A northern blot of LINE-1.3 full-length RNAs produced by the wild-type human LINE-1 element was probed with a strand-specific 5′ UTR probe and probes to either exon 2 (neoExon) or the intron (neoInt) of *neo*^{r}. Bands corresponding to the spliced and unspliced mRNA species are marked FL1.3 (–) intron and FL1.3 (+) intron, respectively. (**c**) Northern blots of endogenous and exogenous LINE-1 RNAs in different species. A northern blot of poly(A)$^+$-selected total LINE-1.3 RNA species produced by the endogenous LINE-1 (EnFL1.3) elements in Ntera2 and HeLa cells and wild-type (WT) and mutant 1 (1 Mut) human LINE-1.3 elements transfected into HeLa and chicken cells was probed with human strand-specific 5′ UTR probe. FL1.3 is full-length human LINE-1.3 mRNA. *neo*^{r} is neomycin-resistance mRNA detected with the randomly labeled probe to exon 2 of the *neo*^{r} gene and was used as a transfection and loading control. β-actin is β-actin mRNA detected by randomly labeled probe to *ACTB* and was used as loading control for endogenous mRNAs in Ntera2 cells.

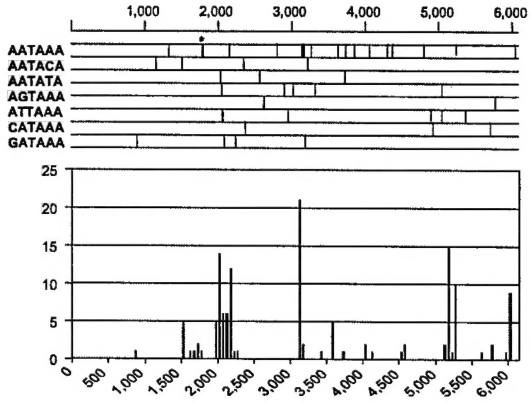

Figure 3 3′ ends of ESTs relative to common LINE-1 poly(A) signals. Map positions are shown for the strongest predicted canonical and noncanonical poly(A) sites in the LINE-1.3 sequence. The asterisk indicates the strongest poly(A) site in our transfection experiments. The chart below shows the relative abundance of human ESTs whose 3′ ends map approximately to the locations shown. This represents 140 ESTs that had been oligo-dT primed and sequenced from the 3′ end. There are only eight full-length LINE-1 elements in this group, representing less than 6% of the total LINE-1-related transcripts. The largest cluster resides around position 2,000, in the region seen near our strongest predicted poly(A) site, as well as those shown in the mutant of that site, and the positions seen in the northern blot of endogenous LINE-1 elements from Ntera2 cells.

(**Supplementary Fig. 3** online) supported the conservation of the process of internal polyadenylation of LINE-1 RNA in mammals. There were few ESTs corresponding to the position of band 3 (**Fig. 2a**) for human LINE-1, but more in the regions of bands 4 and 5 (**Fig. 2a**), consistent with our observations that the relative use of the internal poly(A) sites varied in different cell types. Band 3 may be particularly strong in the cultured cells used in these studies, whereas other sites may be preferred in the endogenous tissues used for most EST studies.

Chicken fibroblasts that lack any endogenous LINE-1 elements but support use of LINE-1 poly(A) signals were transfected with LINE-1.3 (**Fig. 2c**). We used 3′ RACE to amplify the region of the primary truncated LINE-1 RNA products in the northern blots. This analysis showed a pattern consistent with the northern blot results for the wild type and mutant 1 (**Fig. 4a**). There were also multiple larger bands in both elements that corresponded in length to the LINE-1.3 region containing poly(A) signals responsible for the band 4 in the northern-blot assay. This confirms the sites of the functional poly(A) signals and that the relative activity of any given site varies depending on competition with other nearby poly(A) signals[26]. Sequence analysis of eight clones from the strongest 3′ RACE band (**Fig. 4a**) confirmed the use of that site. Analysis of 13 clones from the 2,053–2,153 region of the 3′ RACE gel (**Fig. 4a**) identified 7 clones that terminated at a location consistent with the use of the ATTAAA poly(A) site that was inactivated in mutant 5 (**Fig. 2a**), 3 clones that terminated at a location consistent with the use of the third AATAAA site in LINE-1 (**Fig. 1**), which was inactivated in mutant 3, and 3 clones that used various noncanonical sites located in that general region of LINE-1. All polyadenylations occurred at sites consistent with distance and with the sequence normally associated with 3′ cleavage[26–28]. We carried out 3′ RACE amplification of poly(A)$^+$-selected mRNAs from nontransfected and wild-type LINE-1.3-transfected Ntera2 cells (**Fig. 4b**). The bands were similar in size to those identified in chicken cells. Band 1,788, corresponding to the position of the strongest poly(A) signal, was not detected in the nontransfected Ntera2 cells. It was observed on transfection with the LINE-1.3 expression cassette, however, consistent with differences between endogenous elements in the EST data (**Fig. 3**) and our northern blots of LINE-1.3-transfected cells (**Fig. 2**).

We showed that LINE-1 sequences support the use of a broad range of poly(A) sites that limit retroposition activity of the element, resulting in less retroposition and, therefore, less damage to their hosts' genomes. Despite the low levels of full-length LINE-1 RNA, LINE-1 retrotransposition rates can be quite high, suggesting that the later steps of L1 integration may be very efficient. On the other hand, the internal poly(A) sites may increase the potential damage when LINE-1 elements insert into a gene, causing premature termination of that gene's transcripts. Because of the redundancy in the system, elimination of one or two poly(A) sites may effect only minimally the overall amplification capability of the element. Our observation of a new application of the conventional mRNA 3′ end formation extends the range of the known mechanisms for the control of mRNA levels in mammalian cells and provides additional understanding of the selective pressures that act on new retrotransposition events in reshaping the mammalian genome.

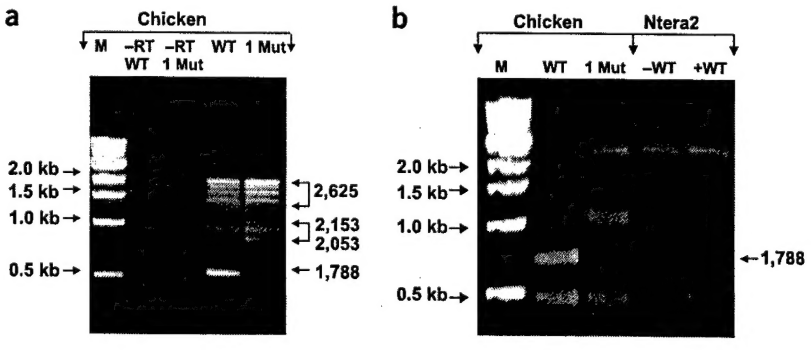

Figure 4 3′ RACE analysis of the prematurely terminated LINE-1.3 RNA species. (**a**) 3′ RACE analysis of poly(A)-selected total mRNAs from chicken fibroblasts transiently transfected with the wild-type (WT) or mutant 1 (1 Mut) LINE-1.3 expression vectors described in **Figure 1**. The upstream primer used in the PCR step corresponds to positions 1,342–1,359 of the LINE-1.3 sense strand. M, 1-kb DNA ladder (NEBiolabs); –RT, without reverse transcriptase. The band corresponding to the strongest predicted poly(A) site identified by the northern blot is labeled 1,788. Genomic location and the hexanucleotide sequence of the identified functional poly(A) sites just upstream of individual bands, after cloning and sequencing, are shown next to their positions on the gel. (**b**) 3′ RACE analysis of poly(A)-selected total mRNAs from chicken fibroblasts transiently transfected with the wild-type (WT) or mutant 1 (1 Mut) LINE-1.3 expression vectors described in **Figure 1** and Ntera2 cells nontransfected (–WT) and transfected with the wild-type LINE-1.3 expression cassette (+WT). The upstream primer used in the PCR step corresponds to positions 1,159–1,188 in the LINE-1.3 sense strand.

METHODS

Cell culture. We maintained NIH 3T3 (ATCC CRL-1658), HeLa (ATCC CCL2) and Ntera2 (ATCC CRL-1973) cells at 37 °C and 5% CO_2 in Dulbecco's modified Eagle medium (GIBCO) high glucose, 10% Colorado calf serum (GIBCO); minimal essential medium (GIBCO), 10% fetal bovine serum (GIBCO); and Dulbecco's modified Eagle medium high glucose (GIBCO), 15% fetal bovine serum, respectively. We maintained chicken fibroblasts (ATCC CRL-12203) at 39 °C and 5% CO_2 in Dulbecco's modified Eagle medium high glucose (GIBCO) and 15% fetal bovine serum.

Transient transfection assay. We transfected 4×10^6 NIH 3T3 cells and 6×10^6 HeLa cells, Ntera2 cells or chicken fibroblasts per 75 cm^2 cell culture flask (Corning) with 3–6 μg of the LINE-1.3 expression cassette by lipofectamine (6–12 μl of Plus reagent, 18–36 μl of lipofectamine; Invitrogen) 16–18 h after plating. We incubated NIH 3T3 cells with the transfection cocktail in the serum-free medium for 4 h and incubated chicken fibroblasts, HeLa cells and Ntera2 cells with the transfection cocktail in the serum-free medium for 3 h. We incubated all cell types in their respective media for 24–26 h before collecting RNA.

RNA extraction and poly(A) selection. We combined the contents of four 75-cm^2 cell culture flasks of each cell type and extracted total mRNA using TRIzol Reagent (Invitrogen). We then carried out chloroform extraction and isopropanol precipitation. We used the PolyATract mRNA isolation system III (Promega) to select poly(A) RNA species as instructed by the manufacturer. We resuspended poly(A)-selected and precipitated RNA in 30 μl of RNase-free water and fractionated it in a single lane of an agarose-formaldehyde gel.

Northern blots. We transferred RNA to a Hybond-N nylon membrane (Amersham Pharmacia Biotech) by capillary transfer overnight at room temperature in a standard 5× sodium chloride/sodium citrate (SSC) solution. We crosslinked the RNA to the membrane with ultraviolet light and prehybridized it in 30% formamide, 1× Denhardt's solution, 1% SDS, 1 M NaCl, 100 μg ml^{-1} salmon sperm DNA and 100 μg ml^{-1} yeast tRNA at 60 °C for at least 6 h. Hybridization with a strand-specific probe (final concentration of $4–8 \times 10^6$ cpm ml^{-1}) was carried out overnight in the same solution at 60 °C. We carried out multiple 10-min washes at high stringency (0.1× SSC, 0.1% SDS) at 60 °C. We quantified the results of the northern-blot assays on a Fuji Phosphorimager.

We generated the strand-specific probe used for the northern-blot assay by the MAXIscript T7 system (Ambion). Primer sequences for generating the template are available on request. We produced DNA template for the probe by PCR with primers that amplified either the LINE-1.3 5′ UTR or exon 2 or the intron of the *neo*^{r} cassette. The T7 promoter sequence was included in the reverse primer of each pair. We fractionated PCR products on a 1% low-melting agarose gel, excised them and purified them using a QIAquick gel extraction kit 50 (QIAGEN).

3′ RACE analysis. We carried out 3′ RACE (Clontech) according to the manufacturer's protocol on poly(A)-selected total mRNA from chicken fibroblasts and Ntera2 cells transiently transfected with LINE-1.3 expression cassette as described above. Primer sequences are available on request. We gel-purified DNA with QIAquick Gel Extraction Kit (QIAGEN), cloned it into pCR2.1-TOPO vector (Invitrogen) and sequenced 3–8 clones for each poly(A) site by automated sequencing.

Site-directed mutagenesis. We used the QuikChange Site-Directed Mutagenesis kit (STRATAGENE) to change the AATAAA hexamer sequence of the strongest predicted poly(A) site in the LINE-1.3 genome (position 1,788) as well as the poly(A) sites at positions 1,323 and 2,154; the poly(A) sites at position 1,323 and 2,154 to GATCAA; the ATTAAA hexamer sequence of the poly(A) signal at position 2,053 to ATCAAG; and the GATAAA hexamer sequence of the poly(A) signal at position 2,079 to GATCAA.

EST searches. We identified LINE-1 ESTs using a BLAST search of the entire human EST database with the LINE-1.3 sequence. We collected the first 1,085 hits, but only 140 corresponded to 3′ sequences. Those sequenced from the 5′ end would not necessarily define the poly(A) site. For **Figure 3**, the 3′ positions of the ESTs were binned into 50-base regions.

Accession numbers. LINE-1.3, L19088; LINE-1Gg-1a, AF036235; LINE-1spa, AF016099.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank C. Schmid and M. Batzer for helpful comments and J. Moran, H. Kazazian Jr. and J. Goodier for providing LINE-1 vectors and for discussion. This work was supported by the US National Institutes of Health (P.D.) and the US Department of Defense Breast Cancer Research Program (V.P.-B.).

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Received 10 September; accepted 27 October 2003
Published online at http://www.nature.com/naturegenetics/

1. Lander, E.S. *et al.* Initial sequencing and analysis of the human genome. International Human Genome Sequencing Consortium. *Nature* **409**, 860–921 (2001).
2. Deininger, P.L., Moran, J.V., Batzer, M.A. & Kazazian, H.H. Jr. Mobile elements and genome evolution. *Curr. Opin. Genet. Dev.* **13**, 1–8 (2003).
3. Ostertag, E.M. & Kazazian, H.H. Jr. Biology of L1 retrotransposons. *Annu. Rev. Genet.* **35**, 501–538 (2001).
4. Kazazian, H.H. Jr. & Moran, J.V. The impact of L1 retrotransposons on the human genome. *Nat. Genet.* **19**, 19–24 (1998).
5. Esnault, C., Maestre, J. & Heidmann, T. Human LINE retrotransposons generate processed pseudogenes. *Nat. Genet.* **24**, 363–367 (2000).
6. Branciforte, D. & Martin, S.L. Developmental and cell type specificity of LINE-1 expression in mouse testis: implications for transposition. *Mol. Cell Biol.* **14**, 2584–2592 (1994).
7. Trelogan, S.A. & Martin, S.L. Tightly regulated, developmentally specific expression of the first open reading frame from LINE-1 during mouse embryogenesis. *Proc. Natl. Acad. Sci. USA* **92**, 1520–1524 (1995).
8. Ostertag, E.M. *et al.* A mouse model of human L1 retrotransposition. *Nat. Genet.* **32**, 655–660 (2002).
9. Bratthauer, G.L., Cardiff, R.D. & Fanning, T.G. Expression of LINE-1 retrotransposons in human breast cancer. *Cancer* **73**, 2333–2336 (1994).
10. Asch, H.L. *et al.* Comparative expression of the LINE-1 p40 protein in human breast carcinomas and normal breast tissues. *Oncol. Res.* **8**, 239–247 (1996).
11. Skowronski, J., Fanning, T.G. & Singer, M.F. Unit-length line-1 transcripts in human teratocarcinoma cells. *Mol. Cell. Biol.* **8**, 1385–1397 (1988).
12. Benihoud, K. *et al.* Unusual expression of LINE-1 transposable element in the MRL autoimmune lymphoproliferative syndrome-prone strain. *Oncogene* **21**, 5593–5600 (2002).
13. Deininger, P.L. & Batzer, M.A. Mammalian retroelements. *Genome Res.* **12**, 1455–1465 (2002).
14. Lerat, E., Capy, P. & Biemont, C. Codon usage by transposable elements and their host genes in five species. *J. Mol. Evol.* **54**, 625–637 (2002).
15. Siebel, C.W. & Rio, D.C. Regulated splicing of the Drosophila P transposable element third intron in vitro: somatic repression. *Science* **248**, 1200–1208 (1990).
16. Medstrand, P., van de Lagemaat, L.N. & Mager, D.L. Retroelement distributions in the human genome: variations associated with age and proximity to genes. *Genome Res.* **12**, 1483–1495 (2002).
17. Tabaska, J.E. & Zhang, M.Q. Detection of polyadenylation signals in human DNA sequences. *Gene* **231**, 77–86 (1999).
18. Moran, J.V., DeBerardinis, R.J. & Kazazian, H.H. Jr. Exon shuffling by L1 retrotransposition. *Science* **283**, 1530–1534 (1999).
19. Dombroski, B.A., Mathias, S.L., Nanthakumar, E., Scott, A.F. & Kazazian, H.H. Isolation of an active human transposable element. *Science* **254**, 1805–1810 (1991).
20. Gilbert, N., Lutz-Prigge, S. & Moran, J.V. Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110**, 315–325 (2002).
21. Mulhardt, C. *et al.* The spastic mouse: aberrant splicing of glycine receptor beta subunit mRNA caused by intronic insertion of L1 element. *Neuron* **13**, 1003–1015 (1994).
22. Naas, T.P. *et al.* An actively retrotransposing, novel subfamily of mouse L1 elements. *EMBO J.* **17**, 590–597 (1998).
23. Sheets, M.D., Ogg, S.C., & Wickens, M.P. Point mutations in AAUAAA and the poly (A) addition site: effects on the accuracy and efficiency of cleavage and polyadenylation in vitro. *Nucleic Acids Res.* **18**, 5799–5805 (1990).
24. Martin, S.L. Ribonucleoprotein particles with LINE-1 RNA in mouse embryonal carcinoma cells. *Mol. Cell. Biol.* **11**, 4804–4807 (1991).
25. Skowronski, J. & Singer, M.F. Expression of a cytoplasmic LINE-1 transcript is regulated in a human teratocarcinoma cell line. *Proc. Natl. Acad. Sci. USA* **82**, 6050–6054 (1985).
26. Zhao, J., Hyman, L. & Moore, C. Formation of mRNA 3′ ends in eukaryotes: mechanism, regulation, and interrelationships with other steps in mRNA synthesis. *Microbiol. Mol. Biol. Rev.* **63**, 405–445 (1999).
27. Colgan, D.F. & Manley, J.L. Mechanism and regulation of mRNA polyadenylation. *Genes Dev.* **11**, 2755–2766 (1997).
28. Proudfoot, N. Poly(A) signals. *Cell* **64**, 671–674 (1991).
29. Moran, J.V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).